



## On theta-palindromic richness

Štěpán Starosta\*

Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2, Czech Republic

### ARTICLE INFO

#### Article history:

Received 6 May 2010

Received in revised form 23 October 2010

Accepted 8 December 2010

Communicated by D. Perrin

#### Keywords:

Combinatorics on words

Pseudopalindrome

Theta-palindrome

Rich words

Theta-episturmian words

### ABSTRACT

In this paper we study generalization of the reversal mapping realized by an arbitrary involutory antimorphism  $\theta$ . It generalizes the notion of a palindrome into a  $\theta$ -palindrome—a word invariant under  $\theta$ . For languages closed under  $\theta$  we give the relation between  $\theta$ -palindromic complexity and factor complexity. We generalize the notion of richness to  $\theta$ -richness and we prove analogous characterizations of words that are  $\theta$ -rich, especially in the case of set of factors invariant under  $\theta$ . A criterion for  $\theta$ -richness of  $\theta$ -episturmian words is given together with other examples of  $\theta$ -rich words.

© 2010 Elsevier B.V. All rights reserved.

### 1. Introduction

A palindrome is a word that reads the same from the left as from the right. In the last decade the study of palindromes in the field of combinatorics on words has notably grown. The motivation comes for instance from the application in mathematical physics or diophantine approximation. Two notions were mainly studied. First the **palindromic complexity** which counts the number of palindromes of given length in an infinite word. Its study resulted for example in a new equivalent definition of Sturmian words using palindromic complexity [12]. The second notion is the notion of **defect** which measures the saturation by palindromic factors of a finite word [6]. An important class of words that are “rich” in palindromes in the utmost sense (called **rich** words) have been defined in [11,13]. They have been further studied and many characterizations have been found. For recent results and summary on rich words see for instance [9,10,13].

A more general approach to the notion of the palindrome has been considered by replacing the reversal mapping by a mapping  $\theta$  having the required properties, i.e., being an involution and antimorphism. A word which is invariant under such mapping is called a **pseudopalindrome** or  $\theta$ -**palindrome**. See for instance [2] or [7]. In this work we concentrate on  $\theta$ -palindromic saturation in infinite words. We derive an upper bound for the number of distinct  $\theta$ -palindromic factors in a finite word. We define an analogue to rich words—words containing the maximum number of  $\theta$ -palindromes possible. For this class of words that are rich in  $\theta$ -palindromes (or shortly  $\theta$ -**rich**) we prove similar equivalent statements as are known for rich words. The main result concerns words with a set of factors closed under  $\theta$ . We prove the following relation between  $\theta$ -palindromic complexity  $\mathcal{P}_\theta(n)$  and factor complexity  $\mathcal{C}(n)$  which is analogous to the case of the reversal mapping [3].  $\mathcal{P}_\theta(n)$  denotes the function counting the number of distinct  $\theta$ -palindromic factors of length  $n$  in a given word and  $\mathcal{C}(n)$  denotes the function counting the number of distinct factors of length  $n$  in a given word.

**Theorem 1.** *Let  $u$  be an infinite word with language closed under  $\theta$ . Then*

$$\mathcal{P}_\theta(n) + \mathcal{P}_\theta(n+1) \leq \mathcal{C}(n+1) - \mathcal{C}(n) + 2, \quad \text{for all } n \in \mathbb{N}.$$

\* Tel.: +420 775131723.

E-mail address: [staroste@jfifi.cvut.cz](mailto:staroste@jfifi.cvut.cz).

We show that the equality is attained for  $\Theta$ -rich words, a result analogous to the case of the reversal mapping in [9].

**Theorem 2.** *Let  $\mathbf{u}$  be an infinite word with language closed under  $\Theta$ . Then  $\mathbf{u}$  is  $\Theta$ -rich if and only if  $\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \mathcal{C}(n + 1) - \mathcal{C}(n) + 2$  for all  $n \geq 1$ .*

We also make a connection to  $\Theta$ -episturmian words that are a generalization of episturmian words—a well known example of rich words. For a survey on such class of words see for instance [8]. Unlike episturmian words,  $\Theta$ -episturmian words are not all  $\Theta$ -rich. We give a condition when this is true.

Section 2 summarizes notation and introduces some needed results from combinatorics on words. It includes known results on rich words and defines  $\Theta$ -palindromes. Section 3 establishes a relation between  $\Theta$ -palindromic and factor complexity. The next section introduces the definition of  $\Theta$ -richness and a proof of Theorem 2. Section 5 provides examples of  $\Theta$ -rich words and also contains examples of  $\Theta$ -episturmian words which are not  $\Theta$ -rich.

## 2. Notation and basic facts

By  $\mathcal{A}$  we denote an *alphabet*—a finite set of symbols, called *letters*. A finite sequence of letters  $w = w_0w_1w_2 \cdots w_{n-1}$  is called a *word*. Its length is denoted by  $|w|$  and equals  $n$ . The set of all words over an alphabet  $\mathcal{A}$ , together with the empty word  $\varepsilon$  and the operation of *concatenation*, form a free monoid  $\mathcal{A}^*$ .

The *reversal* of a finite word  $w = w_0w_1 \cdots w_{n-1}$  is the word  $\bar{w} = w_{n-1}w_{n-2} \cdots w_0$ . The mapping  $w \mapsto \bar{w}$  is called reversal and clearly is a bijection on  $\mathcal{A}^*$ . If  $w = \bar{w}$  then such a word  $w$  is called a *palindrome*.

An infinite sequence  $\mathbf{u} = u_0u_1u_2 \dots$  of letters from  $\mathcal{A}$  is said to be an *infinite word* over  $\mathcal{A}$ . A finite word  $v$  is called a *factor* of a finite word  $w$  if there exist words  $v'$  and  $v''$  such that  $w = v'v''$ . If  $v' = \varepsilon$  then  $v$  is called a *prefix* of  $w$ . If  $v'' = \varepsilon$  then  $v$  is called a *suffix*. A factor of an infinite word  $\mathbf{u}$  is defined analogously with  $v''$  being infinite.

The *language*  $\mathcal{L}(\mathbf{u})$  of an infinite word  $\mathbf{u}$  is the set of all its factors.  $\mathcal{L}_n(\mathbf{u})$  denotes the set of factors of  $\mathbf{u}$  of length  $n$ . We have  $\mathcal{L}_0(\mathbf{u}) = \{\varepsilon\}$  and  $\mathcal{L}(\mathbf{u}) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(\mathbf{u})$ . If for every factor  $w \in \mathcal{L}(\mathbf{u})$  the reversal image of  $w$  is in  $\mathcal{L}(\mathbf{u})$ , then we say that the language  $\mathcal{L}(\mathbf{u})$  is *closed under reversal*.

For any factor  $w \in \mathcal{L}(\mathbf{u})$  there exists an index  $i$  such that  $w = u_iu_{i+1} \cdots u_{i+|w|-1}$ . Such an index is called an *occurrence* of  $w$  in  $\mathbf{u}$ . We will say that a factor  $v$  is *unioccurrent* in some word  $w$  if  $v$  occurs exactly once in  $w$ . If every factor of  $\mathbf{u}$  has at least 2 occurrences in  $\mathbf{u}$  then the word  $\mathbf{u}$  is said to be *recurrent*. One can verify that this is equivalent with every factor occurring infinitely many times. For a recurrent infinite word we define a *complete return word*<sup>1</sup> to a factor  $w \in \mathcal{L}(\mathbf{u})$  to be a factor with exactly 2 distinct occurrences of  $w$ , one as a prefix and one as a suffix.

A factor  $w \in \mathcal{L}(\mathbf{u})$  is said to be *left (right) special* if there exist two distinct letters  $a, b \in \mathcal{A}$  such that  $aw \in \mathcal{L}(\mathbf{u})$  and  $bw \in \mathcal{L}(\mathbf{u})$  ( $wa \in \mathcal{L}(\mathbf{u})$  and  $wb \in \mathcal{L}(\mathbf{u})$ ). A factor is *bispecial* if it is right and left special.

The *factor complexity* of an infinite word  $\mathbf{u}$  is a mapping  $\mathbb{N} \mapsto \mathbb{N}$  defined as  $\mathcal{C}(n) := \#\mathcal{L}_n(\mathbf{u})$ . The *palindromic complexity* is a mapping  $\mathbb{N} \mapsto \mathbb{N}$  defined as  $\mathcal{P}(n) := \#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w = \bar{w}\}$ .

In [3] the following connection between palindromic complexity and the first difference of factor complexity  $\Delta\mathcal{C}(n) = \mathcal{C}(n + 1) - \mathcal{C}(n)$  was proved.

**Theorem 3** ([3]). *Let  $\mathbf{u}$  be an infinite word with language closed under reversal. Then for all  $n$  we have*

$$\mathcal{P}(n) + \mathcal{P}(n + 1) \leq \Delta\mathcal{C}(n) + 2. \tag{1}$$

Given an infinite word  $\mathbf{u}$  and  $n \in \mathbb{N}$ , we define the *Rauzy graph of order  $n$  of the word  $\mathbf{u}$* , denoted  $\Gamma_n(\mathbf{u})$ , as follows.  $\Gamma_n(\mathbf{u})$  is a directed graph whose set of vertices equals  $\mathcal{L}_n(\mathbf{u})$  and set of edges equals  $\mathcal{L}_{n+1}(\mathbf{u})$ . There is an edge  $e \in \mathcal{L}_{n+1}(\mathbf{u})$  going from the vertex  $v_1$  to the vertex  $v_2$  if  $v_1$  is a prefix of  $e$  and  $v_2$  is a suffix of  $e$ . If the word  $\mathbf{u}$  is recurrent then for all  $n$  the graph  $\Gamma_n(\mathbf{u})$  is strongly connected, i.e., there exists a directed path from any vertex to any other vertex. Let  $\deg_+(w)$  denote the outdegree of a vertex  $w$ . Especially for  $\Gamma_n(\mathbf{u})$  we have  $\deg_+(w) = \#\{a \in \mathcal{A} \mid wa \in \mathcal{L}(\mathbf{u})\}$ . Rauzy graphs are a useful tool to calculate the complexity as we have

$$\Delta\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} (\deg_+(w) - 1).$$

An  *$n$ -simple path  $p$*  is a factor of  $\mathbf{u}$  of length at least  $n + 1$  such that its prefix of length  $n$  and its suffix of length  $n$  are both special (left or right) and no interior factor of length  $n$  is special. We can now define the *reduced Rauzy graph of order  $n$  of  $\mathbf{u}$*   $\Gamma'_n(\mathbf{u})$ . Its set of vertices equals the set of special (right or left) factors of  $\mathcal{L}_n(\mathbf{u})$ . There is an edge  $e$  from  $v_1$  to  $v_2$  if there is a  $n$ -simple path  $p$  which begins with  $v_1$  and ends with  $v_2$ . Note that if for a given length  $N$  there is no special factor, the word  $\mathbf{u}$  is periodic, i.e., can be written as  $\mathbf{u} = v^\omega = vvv \dots$ , where  $v$  is a finite word. In such cases the definition of the reduced Rauzy graph of order  $n$  results in an empty graph for all  $n \geq N$ .

Consider now an infinite word  $\mathbf{u}$  with language closed under reversal. We can define the *super reduced Rauzy graph of order  $n$  of  $\mathbf{u}$*  denoted  $\Gamma''_n(\mathbf{u})$  (first defined in [9]). Its set of vertices is formed by unordered pairs  $\{w, \bar{w}\}$ , where  $w \in \mathcal{L}_n(\mathbf{u})$  is a special factor. There is an undirected edge  $\{e, \bar{e}\}$  from  $\{v, \bar{v}\}$  to  $\{w, \bar{w}\}$  if there is an  $n$ -simple path  $e$  starting with  $v$  or  $\bar{v}$  and ending with  $w$  or  $\bar{w}$ . Note that  $\Gamma''_n(\mathbf{u})$  may have multiple edges or loops.

<sup>1</sup> A shorter term “complete return” is also used.

### 2.1. Palindromic richness

We say that a finite word  $w$  is *rich* if it has exactly  $|w| + 1$  palindromic factors. It is in fact the maximum number of palindromic factors possible. An infinite word is rich if all its factors are rich. This definition arises from the work done in [11] and in [13].

The following theorem summarizes some known characterizations of rich words. Listed equivalent definitions of richness have been proven respectively in [13,9,10].

**Theorem 4** ([13,9,10]). *For any infinite word  $\mathbf{u}$  the following conditions are equivalent:*

- (1)  $\mathbf{u}$  is rich,
- (2) any complete return word to a palindromic factor of  $\mathbf{u}$  is a palindrome,
- (3) for any factor  $w$  of  $\mathbf{u}$ , every factor of  $\mathbf{u}$  that contains  $w$  only as its prefix and  $\bar{w}$  only as its suffix is a palindrome,
- (4) each factor of  $\mathbf{u}$  is uniquely determined by its longest palindromic prefix and its longest palindromic suffix.

The following theorem gives a characterization of rich infinite words with language closed under reversal. It also answers the question whether the inequality (1) in Theorem 3 is reached for rich words.

**Theorem 5** ([9]). *Let  $\mathbf{u}$  be an infinite word with language  $\mathcal{L}(\mathbf{u})$  closed under reversal. Then  $\mathbf{u}$  is rich if and only if  $\mathcal{P}(n + 1) + \mathcal{P}(n) = \Delta\mathcal{C}(n) + 2$  for all  $n \in \mathbb{N}$ .*

In [4] a different characterization of rich words with language closed under reversal has been shown. It is based on bispecial orders of factors.

Let us mention two classes of words that we will refer to later and that are also rich. The well-known **Sturmian** words are exactly those words with factor complexity  $\mathcal{C}(n) = n + 1$  for all  $n \in \mathbb{N}$ . The second class consists of so-called **episturmian** words. They are defined as words whose language is closed under reversal and have at most one left special factor of every length.

For further information on episturmian or rich words you can refer for instance to [5,11,13,14]. A differently constructed class of rich words than the two mentioned above may be found in [1].

### 2.2. $\Theta$ -palindromicity

The reversal mapping can in fact be understood as a special case of a more general mapping  $\Theta: \mathcal{A}^* \mapsto \mathcal{A}^*$  which satisfies two conditions:

- (1) it is an involution, i.e.,  $\Theta^2 = \text{id}$ ;
- (2)  $\Theta$  is an antimorphism, i.e.,  $\forall v, w \in \mathcal{A}^*, \Theta(vw) = \Theta(w)\Theta(v)$ .

From now on,  $\Theta$  will denote a mapping with such properties. It is easy to verify that  $\Theta$  can only be a composition of the reversal mapping and an involutory permutation of letters. Such a permutation has two types of cycles distinguished by their length 1 or 2.

The definitions in the previous section can be modified in the following way. A finite word  $w$  is called a  $\Theta$ -palindrome if  $w = \Theta(w)$ .  $\Theta$ -palindromic complexity is a mapping  $\mathbb{N} \mapsto \mathbb{N}$  defined as

$$\mathcal{P}_\Theta(n) := \#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w = \Theta(w)\}.$$

The language of  $\mathbf{u}$  is *closed under  $\Theta$*  if for any factor  $w \in \mathcal{L}(\mathbf{u})$  we have  $\Theta(w) \in \mathcal{L}(\mathbf{u})$ . The definition of the super reduced Rauzy graph is analogous: for words closed under  $\Theta$  the reversal mapping is replaced by  $\Theta$ .

As we will see later, a crucial difference between the reversal mapping and a  $\Theta$  mapping different from the reversal mapping is that there exists a letter  $a$  such that  $a \neq \Theta(a)$ . Therefore there exist words that have no non-empty  $\Theta$ -palindromic suffix at all.

The research on  $\Theta$ -palindromicity has led to several generalizations of known results. We will recall one such generalization. A word is  $\Theta$ -**episturmian** if its language is closed under  $\Theta$  and if for each  $n$  there exists at most one left special factor of length  $n$ . Such words are discussed for instance in [8]. In [2] these words are called pseudopalindromic. A more general class that we will mention later is discussed in [7].

### 3. $\Theta$ -palindromic and factor complexity

As for the reversal mapping, if  $\mathbf{u}$  has its language closed under  $\Theta$ , then it is recurrent. To see that, consider a non- $\Theta$ -palindromic factor  $w \in \mathcal{L}(\mathbf{u})$ . As  $\mathcal{L}(\mathbf{u})$  is closed under  $\Theta$  we can find an occurrence of  $\Theta(w)$  and so find a factor  $r$  having  $w$  as a prefix (respectively suffix) and  $\Theta(w)$  as a suffix (respectively prefix). If  $r$  is not a  $\Theta$ -palindrome, then we can find a different occurrence of  $\Theta(r)$  and in it a second occurrence of  $w$ . If  $w$  or  $r$  is a  $\Theta$ -palindrome it suffices to consider a non- $\Theta$ -palindromic factor that contains  $w$  or  $r$  as factor. Such a factor either exists (for instance one can look only among prefixes of  $\mathbf{u}$ ) or  $\mathbf{u}$  is trivial (periodic with period 1).

The following observation, stated as a lemma, will be used in the proof of Theorem 1. We provide a short proof.

**Lemma 6.** Let  $\mathbf{u}$  be an infinite word with language closed under  $\Theta$ . If there exists an integer  $N$  such that there is no special factor of length  $N$ , then for all  $n \geq N$  we have  $\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = 2$ .

**Proof.** As  $\mathbf{u}$  is recurrent, it is clear that for all  $n \geq N$ ,  $\Gamma_n(\mathbf{u})$  is a cycle. Take  $n \geq N$ . We will consider an undirected graph  $G$  whose vertices are defined by unordered pairs  $\{w, \Theta(w)\}$  for all  $w \in \mathcal{L}_n(\mathbf{u})$ . Let  $e$  be an edge of  $\Gamma_n(\mathbf{u})$  going from the vertex  $w$  to the vertex  $v$ . For each pair  $\{e, \Theta(e)\}$ , there is an edge in  $G$  between the vertex  $\{w, \Theta(w)\}$  and the vertex  $\{v, \Theta(v)\}$ . Note that there may be multiple edges and also loops (in the case  $e = \Theta(e)$ ). According to the terminology already used this graph may be called a “super Rauzy graph of order  $n$ ”.

As  $\Gamma_n(\mathbf{u})$  is a cycle, it is clear that  $G$  is formed by one elementary path  $p$  and at most 2 loops—one possibly on the first vertex of  $p$ , one possibly on the last vertex of  $p$ . Suppose there is an interior vertex  $\{w, \Theta(w)\}$  in  $p$  such that  $w = \Theta(w)$ . Then  $w$  would be a special factor of length  $n$  and we have a contradiction. Denote by  $\{s, \Theta(s)\}$  the first vertex of  $p$ . If there is a loop on  $\{s, \Theta(s)\}$ , then it is clear that  $s \neq \Theta(s)$  as otherwise  $s$  would be special. On the other hand, if there is not a loop on  $\{s, \Theta(s)\}$ , then  $s = \Theta(s)$  as  $\Gamma_n(\mathbf{u})$  would be disconnected. The reasoning is analogous for the last vertex of  $p$ .

As a loop in  $G$  is in fact a  $\Theta$ -palindrome of length  $n + 1$ , we have at most 2  $\Theta$ -palindromes of length  $n$  or  $n + 1$ —one associated with the first vertex of  $p$  and one with the last vertex of  $p$ .  $\square$

We will now prove the analogue of Theorem 3 for languages closed under  $\Theta$ .

**Proof of Theorem 1.** Consider the operation  $\theta$  which to every vertex  $w$  of a Rauzy graph associates  $\Theta(w)$  and to every edge  $e$  associates  $\Theta(e)$ . Since we supposed the language closed under  $\Theta$ ,  $\theta$  is well defined on all Rauzy graphs of  $\mathbf{u}$ . In fact,  $\theta$  maps every Rauzy graph  $\Gamma_n(\mathbf{u})$  onto itself. We will omit  $\mathbf{u}$  in the notation of Rauzy graphs from now on.

Fix  $n$ . We are interested in  $n$ -simple paths. As  $\mathcal{L}(\mathbf{u})$  is closed under  $\Theta$ ,  $\mathbf{u}$  is recurrent. Thus the graph  $\Gamma_n$  is strongly connected and every vertex and every edge belongs to an  $n$ -simple path.

For an edge satisfying  $\theta(e) = e$  we find the  $n$ -simple path  $p$  which contains  $e$ . Then the operation  $\theta$  must map the path represented by  $p$  onto itself, i.e.,  $p = \Theta(p)$ . Similarly, for a non-special vertex  $w$  such that  $\theta(w) = w$ , the  $n$ -simple path containing  $w$  is mapped by  $\theta$  onto itself.

Note that an  $n$ -simple path can contain at most one  $\Theta$ -palindrome of length  $n$  or  $n + 1$  which is not its prefix or suffix. To prove that suppose the contrary. Let  $z$  be an  $n$ -simple path containing two interior  $\Theta$ -palindromic factors of length  $n$  or  $n + 1$ . Denote  $p_1$  the  $\Theta$ -palindrome with leftmost occurrence in  $z$  which is not a prefix of  $z$ . Since there is another  $\Theta$ -palindrome of length  $n$  or  $n + 1$  and  $z$  is a  $\Theta$ -palindrome, there is another occurrence of  $p_1$  as an interior factor of  $z$ . Let  $r$  be the shortest left special factor such that  $p_1$  is its suffix. If there is no occurrence of  $r$  in  $z$ , then there is another occurrence in  $z$  of the prefix of  $z$  of length  $n$  which is neither its prefix nor suffix. If there is at least one occurrence of  $r$  in  $z$ , then either the prefix of length  $n$  of  $r$  is a special factor of  $z$  different from its prefix or suffix, or again the prefix of  $z$  of length  $n$  has another occurrence in  $z$  different from its prefix or suffix. In both cases, we can find an interior occurrence of a special factor of length  $n$ —a contradiction with  $z$  being an  $n$ -simple path.

To give an upper bound on  $\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1)$  therefore consists in finding the number of  $n$ -simple paths in  $\Gamma_n$  that are mapped by  $\theta$  onto themselves and the number of special  $\Theta$ -palindromes of length  $n$ . For that we can consider the reduced Rauzy graph  $\Gamma'_n$ .

The set of vertices of  $\Gamma'_n$  can be partitioned into two disjoint subsets. The first subset is given by special  $\Theta$ -palindromes of length  $n$ . Let  $\alpha$  denote the number of such vertices. The second subset is formed by the rest of the vertices—non- $\Theta$ -palindromic special factors of length  $n$ . Let the number of such vertices be denoted by  $2\beta$ . The total number of vertices in  $\Gamma'_n$  is then  $\alpha + 2\beta$ .

It is clear that the super reduced Rauzy graph  $\Gamma''_n$  has exactly  $\alpha + \beta$  vertices. As  $\Gamma''_n$  is connected, it has at least  $\alpha + \beta - 1$  edges. Thus the number of edges in  $\Gamma'_n$  is at least  $2(\alpha + \beta - 1)$ . These edges correspond to the  $n$ -simple paths in  $\Gamma_n$  which are not mapped by  $\theta$  onto themselves.

As already mentioned, the number of  $\Theta$ -palindromes of length  $n$  or  $n + 1$  is bounded by the number of  $n$ -simple paths in  $\Gamma_n$  which are mapped onto themselves, and the number of special  $\Theta$ -palindromic vertices. We thus have

$$\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) \leq \sum_{w \text{ is special}} \deg_+(w) - 2(\alpha + \beta - 1) + \alpha,$$

where the first summand is the number of all  $n$ -simple paths in  $\Gamma_n$ , the second summand estimates the number of  $n$ -simple paths which are not mapped to themselves, and the third counts the number of special  $\Theta$ -palindromic vertices. We obtain

$$\begin{aligned} \mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) &\leq \sum_{w \in \mathcal{L}_n(\mathbf{u}) \text{ is special}} \deg_+(w) - (\alpha + 2\beta) + 2 \\ &= \sum_{w \in \mathcal{L}_n(\mathbf{u}) \text{ is special}} (\deg_+(w) - 1) + 2 \\ &= \Delta\mathcal{C}(n) + 2. \end{aligned} \tag{2}$$

The proof is finished except for the special case where the word  $\mathbf{u}$  is periodic. This implies the existence of  $N$  such that for all  $n \geq N$  we have  $\Delta\mathcal{C}(n) = 0$ , i.e., there is no special factor of length  $n$ . In this case the reduced Rauzy graph of order  $n$  is empty. According to Lemma 6 we have  $\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = 2$  for all  $n \geq N$ , which completes the proof.  $\square$

**Corollary 7.** Let  $\mathcal{L}(\mathbf{u})$  be closed under  $\Theta$ . Let  $n \geq 0$ . The equality

$$\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \Delta\mathcal{C}(n) + 2 \tag{3}$$

holds if and only if both of the following conditions are satisfied:

- (1) the graph  $\Gamma_n''$  after removing loops is a tree,
- (2) any simple path forming a loop in the graph  $\Gamma_n''$  is mapped by  $\theta$  onto itself.

**Proof.** First, we rewrite the equality in the relation (2) as

$$\Delta\mathcal{C}(n) + 2 = \#n\text{-simple paths in } \Gamma_n' - \alpha - 2\beta + 2$$

where  $\alpha$  is again the number of special  $\Theta$ -palindromes of length  $n$  and  $2\beta$  is the number of non- $\Theta$ -palindromic special factors of length  $n$ . Let  $p$  denote the number of  $n$ -simple paths that are mapped by  $\theta$  onto themselves and  $m$  the number of  $n$ -simple paths that are not mapped onto themselves.

We will prove the equivalence directly. Suppose that  $\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \Delta\mathcal{C}(n) + 2$ . We have

$$\alpha + p = \mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \Delta\mathcal{C}(n) + 2 = p + m - \alpha - 2\beta + 2.$$

Therefore  $m = 2(\alpha + \beta - 1)$ . From the proof of [Theorem 1](#), this is equivalent to having exactly the minimum number of non- $\Theta$ -palindromic  $n$ -simple paths needed so that the graph  $\Gamma_n'$  is strongly connected. As loops are not needed to have the graph strongly connected, they are counted in the number  $p$ . These two facts are equivalent with  $\Gamma_n''$  being a tree after removing loops and all loops being  $\Theta$ -palindromic.  $\square$

A natural question is whether there exist such words for which the equality (3) is satisfied for all  $n$ . The answer is positive only for  $\Theta$  equal to the reversal mapping. In any other case we have  $\mathcal{P}_\Theta(1) < \#\mathcal{A}$  and thus

$$\mathcal{P}_\Theta(0) + \mathcal{P}_\Theta(1) < (\#\mathcal{A} - 1) + 2 = \Delta\mathcal{C}(0) + 2.$$

However, there are words saturated in  $\Theta$ -palindromes that satisfy the equality for all  $n \geq 1$ . Examples of such words are given in [Section 5.1](#).

#### 4. $\Theta$ -Richness

Recall that a finite word is rich if it has the maximum possible number of palindromic factors. If we want to use the same notion of richness for  $\Theta$ -palindromes, we need to take into consideration that there are always fewer  $\Theta$ -palindromes since some letters may not be  $\Theta$ -palindromes. To define richness we need to evaluate the number of these letters contained in a given word. We introduce the set  $\gamma(w)$  as follows

$$\gamma(w) := \{ \{a, \Theta(a)\} \mid a \in \mathcal{A}, a \neq \Theta(a), a \text{ or } \Theta(a) \in \mathcal{L}(w) \}.$$

It is clear that if  $w'$  is a prefix of some finite word  $w$ , then  $\#\gamma(w') \leq \#\gamma(w)$ . It is also clear that if for some letter  $x \in \mathcal{A}$  we have  $\gamma(w) \neq \gamma(wx)$ , then the word  $wx$  has no  $\Theta$ -palindromic suffix as  $\Theta(x)$  does not occur in it. Also, as  $\gamma(wx) = \gamma(w) \cup \{ \{x, \Theta(x)\} \}$ , one can see that in this case  $\#\gamma(wx) = \#\gamma(w) + 1$ .

The following proposition gives the maximum number of  $\Theta$ -palindromes that may occur in a finite word.

**Proposition 8.** Let  $w$  be a finite word. Then the number of its  $\Theta$ -palindromic factors is bounded by  $|w| + 1 - \#\gamma(w)$ .

**Proof.** The proof is done by induction on  $|w|$ . The statement is clear for  $|w| \leq 1$ . Suppose that for  $|w| \leq N$ ,  $w$  contains at most  $|w| + 1 - \#\gamma(w)$  different  $\Theta$ -palindromes. Let  $x \in \mathcal{A}$ . We will count the number of  $\Theta$ -palindromes in the word  $wx$ . It is clear that the number of  $\Theta$ -palindromes in  $wx$  is equal to the number of  $\Theta$ -palindromes in  $w$  plus the number of new  $\Theta$ -palindromic factors— $\Theta$ -palindromic suffixes of  $wx$  that do not occur in  $w$ .

First suppose we have  $\gamma(w) \neq \gamma(wx)$ . As there is no  $\Theta$ -palindromic suffix the number of  $\Theta$ -palindromes in  $wx$  equals the number of  $\Theta$ -palindromes in  $w$  which is at most  $|w| + 1 - \#\gamma(w) = |wx| + 1 - \#\gamma(wx)$ .

Suppose  $\gamma(wx) = \gamma(w)$ . Suppose for contradiction that  $wx$  contains more than  $|wx| + 1 - \#\gamma(wx)$  different  $\Theta$ -palindromes. As  $|wx| + 1 - \#\gamma(wx) = (|w| + 1 - \#\gamma(w)) + 1$ , it is clear that there must be at least two different  $\Theta$ -palindromes  $p$  and  $q$  such that  $w = yp$  and  $w = zq$  that do not occur previously in  $w$  (as they were not yet counted). We can assume that  $|p| > |q|$  and  $p$  is the longest of such suffixes of  $wx$ . On the other hand we have  $p = vq = q\Theta(v)$  where  $|v| > 0$ . Therefore every such factor  $q$  occurs already in  $w$  and has already been counted in the number of  $\Theta$ -palindromes in  $w$ —a contradiction. We can see that the number of  $\Theta$ -palindromes in  $wx$  is equal to either the number of  $\Theta$ -palindromes in  $w$  or the number of  $\Theta$ -palindromes in  $w$  plus 1. Therefore the upper bound is  $|w| + 1 - \#\gamma(w) + 1$  which equals  $|wx| + 1 - \#\gamma(wx)$ .  $\square$

We can see from the preceding proof that if there is a new  $\Theta$ -palindromic factor, it is the longest  $\Theta$ -palindromic suffix. We will now define a class of words having maximum number of  $\Theta$ -palindromes.

**Definition 9.** A finite word  $w$  is  $\Theta$ -rich if it contains  $|w| + 1 - \#\gamma(w)$   $\Theta$ -palindromic factors.

An infinite word is  $\Theta$ -rich if all its factors are  $\Theta$ -rich.

One can see that all factors of a finite  $\Theta$ -rich word are  $\Theta$ -rich as well. Thus, if an infinite word has all its prefixes  $\Theta$ -rich, then all its factors are  $\Theta$ -rich and it is by definition  $\Theta$ -rich.

The following equivalence is a consequence of the last remark and of the proof of Proposition 8. It is a  $\Theta$ -rich analogue of results in [11,13].

**Proposition 10.** *An infinite or finite word  $\mathbf{u}$  is  $\Theta$ -rich if and only if the longest  $\Theta$ -palindromic suffix of each prefix  $p$  is unioccurrent except for prefixes having the form  $px$ , with  $x \in \mathcal{A}$  and  $\gamma(p) \neq \gamma(px)$ .*

Proposition 10 leads to the following corollary, which is a  $\Theta$ -rich analogue of a result in [9].

**Corollary 11.** *Let  $\mathbf{u}$  be an infinite  $\Theta$ -rich word. Then for all  $w \in \mathcal{L}(\mathbf{u})$ ,  $w \neq \Theta(w)$ , the occurrences of  $w$  and  $\Theta(w)$  alternate.*

**Proof.** Suppose there exist a factor  $w \in \mathcal{L}(\mathbf{u})$ ,  $w \neq \Theta(w)$ , and a factor  $v \in \mathcal{L}(\mathbf{u})$ ,  $|v| > |w|$ , such that  $w$  is its prefix and its suffix and there is no occurrence of  $\Theta(w)$  in  $v$ . Since the last letter of  $v$  already occurred in  $v$  and  $v$  does not have a unioccurrent  $\Theta$ -palindromic suffix, according to Proposition 10 we have a contradiction.  $\square$

The following proposition gives an equivalent characterization of  $\Theta$ -rich words.

**Proposition 12.** *Let  $\mathbf{u}$  be an infinite or finite word. Then  $\mathbf{u}$  is rich  $\Theta$ -rich if and only if for each factor  $v \neq \varepsilon$  of  $\mathbf{u}$ , any factor of  $\mathbf{u}$  beginning with  $v$  and ending with  $\Theta(v)$ , containing no other occurrences of  $v$  or  $\Theta(v)$ , is a  $\Theta$ -palindrome; and for all  $a \in \mathcal{A}$ , the occurrences of  $a$  and  $\Theta(a)$  in  $\mathbf{u}$  alternate.*

**Proof.** ( $\Rightarrow$ ): Let  $v$  be a factor of  $\mathbf{u}$ . Consider a factor  $r$  containing  $v$  as a prefix and  $\Theta(v)$  as a suffix, with no other occurrences of  $v$  or  $\Theta(v)$ .

If  $|v| = |r|$ , then  $v = \Theta(v)$  and  $r$  is a  $\Theta$ -palindrome.

Let us assume  $|v| < |r|$ . Let  $s$  be the longest  $\Theta$ -palindromic suffix of  $r$ . (There is such a suffix as  $r$  is  $\Theta$ -rich and the image by  $\Theta$  of the last letter of  $r$  has already occurred in  $r$  as the first letter.)

If  $|s| \leq |v|$  then we can write  $\Theta(v) = zs$  with  $z \in \mathcal{A}^*$ . Then  $v = \Theta(s)\Theta(z) = s\Theta(z)$ . We have a second occurrence of  $s$  in  $r$ —a contradiction with its unioccurrence.

If  $|v| < |s| < |r|$  then as  $\Theta(v)$  is a suffix of  $s$  we can write  $s = z\Theta(v) = v\Theta(z)$ . Therefore we can find a second occurrence of  $v$  in  $r$ —a contradiction with the choice of  $r$ .

The last possibility left is  $s = r$  and thus  $r$  is a  $\Theta$ -palindrome.

Alternation of  $a$  and  $\Theta(a)$  follows from Corollary 11.

( $\Leftarrow$ ): Suppose for contradiction that  $\mathbf{u}$  is not  $\Theta$ -rich. Then there is a prefix  $p$  whose longest  $\Theta$ -palindromic suffix  $s$  occurs at least twice in  $p$  or there is a prefix  $p$  with no non-empty  $\Theta$ -palindromic suffix such that  $p$  ends in a non- $\Theta$ -palindromic letter  $a$  and has at least one another occurrence of  $a$  or  $\Theta(a)$ .

In the first case, as  $s$  occurs in  $p$  for a second time, there is a complete return word to  $s$  contained in  $p$  which is its suffix. The complete return word is a  $\Theta$ -palindrome and is longer than  $s$ —a contradiction.

In the second case, as occurrences of  $a$  and  $\Theta(a)$  alternate, we can find the rightmost occurrence of  $\Theta(a)$  in  $p$ . The suffix starting at that occurrence is then a word beginning with  $\Theta(a)$ , ending with  $a$  and with no other occurrences of  $a$  or  $\Theta(a)$ . We supposed such words were  $\Theta$ -palindromic. Thus  $p$  has a  $\Theta$ -palindromic suffix and we have a contradiction.  $\square$

In the last proof we did not need all the assumptions to prove the second implication. We can formulate a weaker sufficient condition as follows.

**Proposition 13.** *Let  $\mathbf{u}$  be an infinite or finite word satisfying*

- all complete return words to  $\Theta$ -palindromes are  $\Theta$ -palindromes;
- for any  $a \in \mathcal{A}$ ,  $a \neq \Theta(a)$ , occurrences of  $a$  and  $\Theta(a)$  alternate;
- factors beginning with  $a \in \mathcal{A}$  and ending with  $\Theta(a)$ , with no other occurrences of  $a$  or  $\Theta(a)$ , are  $\Theta$ -palindromes.

Then  $\mathbf{u}$  is  $\Theta$ -rich.

The following lemma will be used in the proof of Theorem 2.

**Lemma 14.** *Let  $\mathbf{u}$  be an infinite word with language closed under  $\Theta$ . If  $\mathcal{P}_\Theta(1) + \mathcal{P}_\Theta(2) = \Delta\mathcal{C}(1) + 2$  then for all letters  $a$ ,  $a \neq \Theta(a)$ , occurrences of  $a$  and  $\Theta(a)$  alternate.*

**Proof.** Let  $a \in \mathcal{A}$  and suppose  $a$  is not a  $\Theta$ -palindrome. We will look at the super reduced Rauzy graph  $\Gamma_1''$  for which the statements in Corollary 7 hold.

Suppose  $a$  is not special. Then  $a$  is a factor of exactly one 1-simple path  $p$ . If the edge  $\{p, \Theta(p)\}$  is a loop in  $\Gamma_1''$  then the occurrences of  $a$  and  $\Theta(a)$  must alternate as  $p = \Theta(p)$ . If  $\{p, \Theta(p)\}$  is not a loop then there is no path in  $\Gamma_1''$  that would contain 2 occurrences of  $p$  and no occurrence of  $\Theta(p)$ . Such a path would form a cycle in  $\Gamma_1''$  which is not possible.

Suppose now that  $a$  is special. Then  $\{a, \Theta(a)\}$  is a vertex of  $\Gamma_1''$ . We will look at paths going from  $\{a, \Theta(a)\}$  again to  $\{a, \Theta(a)\}$  which are not going through the vertex  $\{a, \Theta(a)\}$ . There are 2 possibilities.

In the first case such path corresponds to a loop in  $\Gamma_1''$  connected to  $\{a, \Theta(a)\}$ . Therefore, the corresponding path in  $\Gamma_1$  is a 1-simple path and is a  $\Theta$ -palindrome beginning with  $a$  or  $\Theta(a)$  respectively and ending with  $\Theta(a)$  or  $a$  respectively.

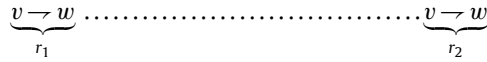


Fig. 1. Two consecutive realizations  $v \rightarrow w$ .

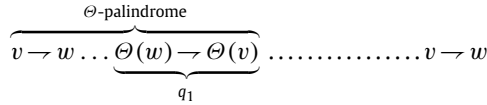


Fig. 2. The leftmost occurrence of realization  $\Theta(w) \rightarrow \Theta(v)$ .

In the second case such path starts with an edge  $\{e, \Theta(e)\}$ . It is clear that the end of this path is again the same edge  $\{e, \Theta(e)\}$  as otherwise there would be a cycle in  $\Gamma''_1$ . If the corresponding path in  $\Gamma''_1$  starts with  $e$ , it must end with  $\Theta(e)$  otherwise there would be a multiple edge (forming a cycle) between the vertices of  $\Gamma''$  connected by  $\{e, \Theta(e)\}$ . Therefore, if such a path starts with  $a$ , it must end with  $\Theta(a)$  and the proof is finished.  $\square$

In what follows, to ease the notation, we will borrow some elements from the free group generated by  $\mathcal{A}$  and use them in the following way. Let  $w$  be a finite word. By  $w^{-1}$  we denote an element whose concatenation with  $w$  produces the empty word, i.e.,  $ww^{-1} = w^{-1}w = \varepsilon$ . The notation is justified by the fact that we will never concatenate  $w^{-1}$  with anything other than  $w$ .

**Proof of Theorem 2.**

( $\Leftarrow$ ):

Suppose that  $\mathbf{u}$  is not rich. According to Lemma 14 non- $\Theta$ -palindromic letters alternate. Proposition 12 implies there exists a factor  $v \in \mathcal{L}(\mathbf{u})$  and a factor  $r \in \mathcal{L}(\mathbf{u})$  such that  $r$  begins with  $v$  and ends with  $\Theta(v)$ , with no other occurrences of  $v$  or  $\Theta(v)$ , and  $r$  is not a  $\Theta$ -palindrome. Then  $r$  is of the form

$$r = vwxsy\Theta(w)\Theta(v),$$

where  $w$  and  $s$  are finite words and  $x, y \in \mathcal{A}, x \neq \Theta(y)$ . As the language is closed under  $\Theta$ , we can see that  $vw\Theta(y) \in \mathcal{L}(\mathbf{u})$ . This means that the factor  $vw$  is right special and therefore  $\{vw, \Theta(vw)\}$  is a vertex in  $\Gamma''_{|vw|}$ .

If  $r$  is a  $|vw|$ -simple path then  $r$  forms a loop in  $\Gamma''_{|vw|}$ . Therefore according to Corollary 7 it is a  $\Theta$ -palindrome—a contradiction.

Suppose now that  $r$  is not a  $|vw|$ -simple path, i.e., contains other special factors of length  $|vw|$ . Denote by  $p, q$  two factors of  $r$  such that they are  $|vw|$ -simple paths and  $p$  is a prefix and  $q$  a suffix of  $r$ . Let  $p'$  denote the special suffix of length  $|vw|$  of  $p$  and  $q'$  the special prefix of  $q$  of length  $|vw|$ . It is clear that vertices  $\{vw, \Theta(vw)\}$  and  $\{p', \Theta(p')\}$  are connected as well as  $\{vw, \Theta(vw)\}$  and  $\{q', \Theta(q')\}$  are connected.

If  $p'$  and  $q'$  coincide, i.e.,  $r = pp'^{-1}q$ , then as  $x \neq \Theta(y)$  we have a multiple edge in  $\Gamma''_{|vw|}$ —a cycle in  $\Gamma''_{|vw|}$  and thus a contradiction. If  $p'$  and  $q'$  do not coincide there exists a path in  $\Gamma''_{|vw|}$  leading from  $\{p', \Theta(p')\}$  to  $\{q', \Theta(q')\}$ , avoiding  $\{vw, \Theta(vw)\}$ , which together with the two paths connecting  $\{vw, \Theta(vw)\}$  form a cycle and we have again a contradiction with Corollary 7.

( $\Rightarrow$ ):

We say that a factor  $r$  is a realization of an edge in  $\Gamma''$  between vertices  $\{v, \Theta(v)\}$  and  $\{w, \Theta(w)\}$  if

- $r$  has a prefix  $v$  or  $\Theta(v)$  and a suffix  $w$  or  $\Theta(w)$  OR
- $r$  has a prefix  $w$  or  $\Theta(w)$  and a suffix  $v$  or  $\Theta(v)$ ,

and contains no interior occurrences of factors  $v, \Theta(v), w$  and  $\Theta(w)$ .

We will denote by  $v \rightarrow w$  a realization of an edge between vertices  $\{v, \Theta(v)\}$  and  $\{w, \Theta(w)\}$  which starts with  $v$  and ends with  $w$ .

First, we are going to show that for every pair of distinct vertices  $\{v, \Theta(v)\}$  and  $\{w, \Theta(w)\}$  there are exactly two realizations of an edge between them and that they are  $\Theta$  images of each other. Let  $w \neq v$  and  $w \neq \Theta(v)$ . Suppose without loss of generality there is a realization  $v \rightarrow w$ . We will find two consecutive realizations, denoted  $r_1$  and  $r_2$ . See Fig. 1.

According to Corollary 11 occurrences of each factor alternate with occurrences of its  $\Theta$  image. Therefore we can find occurrences of  $\Theta(v)$  and  $\Theta(w)$  between occurrences of  $r_1$  and  $r_2$ . Let us first find the leftmost occurrence (after  $r_1$ ) of  $\Theta(v)$ . As every factor beginning with  $v$  and ending with  $\Theta(v)$  is a  $\Theta$ -palindrome, we can find a realization  $\Theta(w) \rightarrow \Theta(v)$ . Denote this realization by  $q_1$ . See Fig. 2. Note that the occurrences of the factor  $w$  in  $r_1$  and  $\Theta(w)$  in  $q_1$  may coincide.

Let us now find the rightmost occurrence (before  $r_2$ ) of  $\Theta(w)$ . Analogously, we find a realization  $\Theta(w) \rightarrow \Theta(v)$ . Denote this realization by  $q_2$ . See Fig. 3.

Suppose now that occurrences of  $q_1$  and  $q_2$  are not equal. If they are not consecutive, we can find a realization of  $\Theta(w) \rightarrow \Theta(v)$ , denoted by  $q'_1$ , such that  $q'_1$  and  $q_2$  are consecutive. Now, using the same arguments as for consecutive

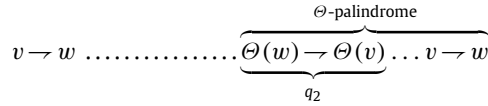


Fig. 3. The rightmost occurrence of realization  $\Theta(w) \rightarrow \Theta(v)$ .

realizations  $r_1$  and  $r_2$ , we can find a realization  $v \rightarrow w$  between  $q'_1$  and  $q_2$ . This is a contradiction with  $r_1$  and  $r_2$  being consecutive. Thus the occurrence of  $q_1$  equals the occurrence  $q_2$  and  $q_1 = q_2$ .  $\Theta$ -palindromes depicted in Figs. 2 and 3 imply  $q_1 = \Theta(r_1)$  and  $q_2 = \Theta(r_2)$ , respectively. Therefore  $r_1 = r_2$  and we deduce that every realization  $v \rightarrow w$  equals  $r_1$  and every realization  $\Theta(v) \rightarrow \Theta(w)$  equals  $\Theta(r_1)$ .

So far we have shown that for every pair of distinct vertices  $\{v, \Theta(v)\}$  and  $\{w, \Theta(w)\}$  there are exactly two realizations of an edge between them and that they are  $\Theta$  images of each other. This implies that  $\Gamma''$  after removing loops is a tree. Corollary 11 claims that occurrences of  $w$  and  $\Theta(w)$  alternate, thus any realization between the same vertices  $\{w, \Theta(w)\}$  is of the form  $w \rightarrow \Theta(w)$  or  $\Theta(w) \rightarrow w$ . By Proposition 12, such realizations are  $\Theta$ -palindromes. These realizations are in fact loops in  $\Gamma''$ . Using Corollary 7 these two parts together imply that the equality (3) holds.  $\square$

**5. Examples of  $\Theta$ -rich words**

In this section we will provide examples of  $\Theta$ -rich words. We first concentrate on a generalization of episturmian words, namely  $\Theta$ -episturmian words, mentioned earlier. The following definition will serve to reveal the structure of the language of a  $\Theta$ -episturmian word. Its main aim is to ease the notation.

**Definition 15.** Let  $\mathbf{u}$  be a  $\Theta$ -episturmian word. Denote by  $(w_k)_{k=0}^K$  the sequence of all bispecial factors of  $\mathbf{u}$ , ordered by increasing length and starting by  $w_0 = \varepsilon$  with  $K$  possibly equal to  $+\infty$ . Furthermore let  $z_k$  denote the unique letter such that  $w_k z_k$  is a prefix of  $w_{k+1}$ . Denote by  $p_k \in \mathcal{A}^*$  the word such that  $w_{k+1} = p_k \Theta(z_k) w_k$ . For all  $k \in \mathbb{N}$  and all  $a \in \mathcal{A}$  let  $r_{k,a}$  denote the complete return word to  $w_k$  beginning with  $w_k a$ , if it exists.

For a letter  $a \neq \Theta(a)$ , let  $k_a$  denote the maximal index  $k \leq K$ , if it exists, such that  $w_k$  has non- $\Theta$ -palindromic complete return words beginning with  $w_k a$  and with  $w_k \Theta(a)$ . If  $k_a$  does not exist, we will consider it to be  $+\infty$ .

Note that since there is at most one complete return word  $w_k$  having a prefix  $w_k a$ , the definition of  $r_{k,a}$  is correct. As the language of  $\mathbf{u}$  is closed under  $\Theta$ ,  $r_{k,a}$  exists if and only if  $r_{k,\Theta(a)}$  exists. Note also that  $k_a \geq 1$  for every letter  $a \neq \Theta(a)$ .

**Proposition 16.** Let  $\mathbf{u}$  be a  $\Theta$ -episturmian word.

- (1) If  $a = \Theta(a)$  and  $r_{k,a}$  exists, then  $r_{k,a}$  is a  $\Theta$ -palindrome;
- (2) if  $a \neq \Theta(a)$ , then
  - a. if  $k \leq k_a$ , then  $r_{k,a}$  equals  $w_k a w_k$ , furthermore  $a$  is not contained in  $w_k$  or in any other complete return word to  $w_k$  except for  $r_{k,a}$ ;
  - b. if  $k > k_a$ , then at most one of the words  $r_{k,a}$  and  $r_{k,\Theta(a)}$  exist.

**Proof.** The proof will be done by induction on  $k$ .

For  $k = 0$ , we have  $w_0 = \varepsilon$  and the claim holds trivially (for non- $\Theta$ -palindromic  $a$  we are in the case (2a)).

Suppose the claim holds for  $k$ . We will prove it for  $k + 1$ . We need to consider especially the case  $k + 1 = k_a + 1$  for all  $a$  to be able to assume also (2b).

Case (1):

Let  $a = \Theta(a)$ . If  $a = z_k$  then  $r_{k+1,a} = p_k \Theta(z_k) w_k a s$  for some word  $s$ . As there is at most one complete return word to  $w_k$  beginning with  $w_k a$ , we have  $w_k a s = w_k a \Theta(p_k)$ . Thus,  $r_{k+1,a} = p_k a w_k a \Theta(p_k)$  which is a  $\Theta$ -palindrome.

If  $a \neq z_k$  suppose that  $r_{k+1,a}$  is not a  $\Theta$ -palindrome. Therefore we can write  $r_{k+1,a} = p_k \Theta(z_k) w_k a x$  and  $w_k a x = y a w_k v = r_{k,a} v$  with  $x, y, v \in \mathcal{A}^*$ . By the assumption  $r_{k,a}$  is a  $\Theta$ -palindrome which implies that  $v$  cannot start with  $z_k$  as  $r_{k+1,a}$  would be a  $\Theta$ -palindrome. In other words we can find a factor having the form  $a w_k b$  with  $b \neq z_k$ .

We can then find two different factors  $\Theta(z_k) w_k a \in \mathcal{L}(\mathbf{u})$  and  $\Theta(b) w_k a \in \mathcal{L}(\mathbf{u})$ . Thus  $w_k a$  is left special which is a contradiction with  $w_k z_k$  being a second left special factor of the same length. Thus  $b = z_k$  and we have  $r_{k+1,a} = p_k \Theta(z_k) r_{k,a} z_k \Theta(p_k)$  which is a  $\Theta$ -palindrome. This concludes the proof of the case (1).

Case (2):

Let now  $a \neq \Theta(a)$ . We distinguish the following three cases according to the value of  $k_a$ .

I.  $k + 1 \leq k_a$

Suppose that  $a$  is contained in  $w_{k+1}$ . This means that  $w_k a w_k$  is a factor of  $w_{k+1}$  and so is  $w_k \Theta(a) w_k$ .

Suppose  $z_k = a$ . We have  $w_{k+1} = w_k a w_k p w_k \Theta(a) w_k$  where  $p$  is a  $\Theta$ -palindrome or  $w_k^{-1}$ . Then the complete return word  $r_{k+1,a}$  equals  $w_{k+1} a w_k p w_k \Theta(a) w_k$  which is a  $\Theta$ -palindrome and contradicts the assumption  $k + 1 < k_a$ .

Suppose now  $z_k \neq a$ . We can then find a factor of  $w_{k+1}$  of the form  $b w_k a w_k c$  with  $b \neq \Theta(z_k)$ . (If  $b = \Theta(z_k)$ , then we would have a shorter bispecial factor than  $w_{k+1}$  but longer than  $w_k$ .) Since  $r_{k+1,a}$  exists we can find a factor  $\Theta(z_k) w_k a$ . Thus  $w_k a$  is right special which contradicts  $w_k z_k$  being right special of the same length.

We will now prove that  $r_{k+1,a}$  is the only complete return word to  $w_{k+1}$  to contain  $a$ . Suppose that  $a$  is contained in  $r_{k+1,b}$  with  $b \neq a$ . Since  $a$  is not contained in  $w_{k+1}$ ,  $r_{k+1,b}$  is of the form  $w_{k+1} x w_{k+1}$  where  $x \in \mathcal{A}^*$  contains  $a$ .



We can find a factor of  $x$  equal to  $cw_k a$  with  $c \neq \Theta(z_k)$ . As  $r_{k+1,a}$  exists there is a factor  $\Theta(z_k)w_k a$  which implies that  $w_k a$  is right special. Since especially  $a \neq z_k$  as otherwise it would be contained in  $w_{k+1}$ , we arrive at a contradiction.

It remains to show that  $r_{k+1,a} = w_{k+1} a w_{k+1}$ . Since  $k + 1 < k_a$  we know from the definition of  $k_a$  that both  $r_{k+1,a}$  and  $r_{k+1,\Theta(a)}$  exist. We can see that  $r_{k+1,a} = p_k \Theta(z_k) r_{k,a} x$  with  $x \in \mathcal{A}^*$ . If the first letter of  $x$  is  $z_k$  we are finished. Suppose it is not true and denote that letter by  $x_0 \neq z_k$ . We have  $aw_k x_0 \in \mathcal{L}(\mathbf{u})$ . As  $r_{k+1,\Theta(a)}$  exists (from the definition of  $k_a$ ) we know that  $\Theta(z_k)w_k \Theta(a) \in \mathcal{L}(\mathbf{u})$  which implies  $aw_k z_k \in \mathcal{L}(\mathbf{u})$ . This implies that  $aw_k$  is right special. As  $\Theta(z_k) \neq a$  we can see that we have 2 right special factors of the same length and thus  $x_0$  must be  $z_k$ . Finally as  $r_{k,a} = w_k a w_k$  we have  $r_{k+1,a} = p_k \Theta(z_k) r_{k,a} z_k \Theta(p_k) = w_{k+1} a w_{k+1}$ .

II.  $k + 1 = k_a + 1$

If  $w_{k+1}$  does not exist, then  $r_{k+1,a}$  does not exist too.

If  $w_{k+1}$  exists, suppose that both  $r_{k+1,a}$  and  $r_{k+1,\Theta(a)}$  exist and at least one of them is a  $\Theta$ -palindrome. We can suppose without loss of generality that  $r_{k,a}$  is a  $\Theta$ -palindrome. We can then find a factor  $\Theta(a)w_k z_k$ .  $r_{k+1,\Theta(a)}$  contains a factor  $\Theta(a)w_k b$  with  $b \neq z_k$  as otherwise we would have two complete return words  $r_{k+1,a}$ . The factor  $\Theta(a)w_k$  is therefore right special which implies  $z_k = a$ . As  $\Theta(a)w_k \Theta(a) \in \mathcal{L}(\mathbf{u})$  we have also  $aw_k a \in \mathcal{L}(\mathbf{u})$  and  $aw_k$  is not right special. Since  $w_k a$  can only be extended as  $w_k a w_k$  we arrive to a contradiction— $\mathbf{u} = pw_k a w_k a w_k a \dots$  which implies  $\mathbf{u}$  is not recurrent ( $\Theta(a)$  occurs only in the word  $p$ ).

III.  $k + 1 > k_a + 1$

It suffices to see that if  $r_{k,a}$  does not exist then  $r_{k+1,a}$  does not exist too.  $\square$

**Corollary 17.** For all letters  $a \neq \Theta(a)$  the value of  $k_a$  is finite.

**Proof.** If  $\mathbf{u}$  is periodic, then  $K$  is finite and so is  $k_a$ . If  $\mathbf{u}$  is aperiodic and  $k_a = +\infty$  then we stay in the case (2a) for all  $k$ . In other words all bispecial factors do not contain  $a$  which is clearly impossible.  $\square$

From the above corollary we can deduce that if  $\Theta$  is different from the reversal mapping, then there exists  $k$  such that  $\Delta\mathcal{C}(|w_k|) < \#\mathcal{A} - 1$ , i.e., the growth of factor complexity is not constant.

The following corollary will serve to derive the criterion for richness of  $\Theta$ -episturmian words.

**Corollary 18.** Let  $k > k_a$  and  $r_{k,a}$  is a  $\Theta$ -palindrome. Then either  $r_{k+1,a}$  does not exist or is a  $\Theta$ -palindrome.

**Proof.** Suppose that  $r_{k+1,a}$  exists and is not a  $\Theta$ -palindrome. Then  $r_{k+1,a}$  begins by  $p_k \Theta(z_k) r_{k,a} b$  where  $z_k \neq b \in \mathcal{A}$ . As  $\Theta(a)w_k b \in \mathcal{L}(\mathbf{u})$  so is  $\Theta(b)w_k a \in \mathcal{L}(\mathbf{u})$ . Together with  $\Theta(z_k)w_k a \in \mathcal{L}(\mathbf{u})$  we see that  $w_k a$  is left special, i.e.,  $z_k = a$ . This is impossible as we would have  $r_{k+1,a} = p_k \Theta(a)w_k a \Theta(p_k)$ —a  $\Theta$ -palindrome.  $\square$

**Corollary 19.** Let  $\mathbf{u}$  be a  $\Theta$ -episturmian word. Let  $l_k = |w_k|$ . If there exists an index  $k_0$  such that  $\mathcal{P}_\Theta(l_{k_0}) + \mathcal{P}_\Theta(l_{k_0} + 1) = \Delta\mathcal{C}(l_{k_0}) + 2$ , then for all  $n \geq l_{k_0}$

$$\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \Delta\mathcal{C}(n) + 2.$$

**Proof.** First note that as  $\mathbf{u}$  is  $\Theta$ -episturmian,  $\Gamma_n''$  has at most one vertex for all  $n$ .

According to Corollary 7, the graph  $\Gamma_{l_{k_0}}''$  consists of one vertex and  $\mathcal{P}_\Theta(l_{k_0}) + \mathcal{P}_\Theta(l_{k_0} + 1) - 1$  edges which are loops mapped onto themselves. Therefore all complete return words to  $w_{k_0}$  are  $\Theta$ -palindromes. Thus, using Corollary 18, for all  $k > k_0$ , the graph  $\Gamma_k''$  consists of one vertex with loops mapped onto themselves. Thus using again Corollary 7, we see that  $\mathcal{P}_\Theta(l_k) + \mathcal{P}_\Theta(l_k + 1) = \Delta\mathcal{C}(l_k) + 2$ .

We will prove that for all  $k > k_0$ , for all  $n$  such that  $l_{k-1} < n < l_k$ , the graph  $\Gamma_n''$  is of the same form as  $\Gamma_k''$  (one vertex with loops mapped onto themselves) and thus according to Corollary 7 the equation  $\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \Delta\mathcal{C}(n) + 2$  holds. It suffices to see that every  $n$ -simple path is in fact a  $\Theta$ -palindromic factor of a  $(l_k)$ -simple path or  $w_k$ .

It remains to see the case of the finite value of  $K$ . Then for all  $n > K$ , there is no special factor of length  $n$ . According to Lemma 6, we have  $2 = \mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \Delta\mathcal{C}(n) + 2$ .  $\square$

The following corollary stems from combining the previous claim with Theorem 2. It gives a criterion to see whether a  $\Theta$ -episturmian word is  $\Theta$ -rich or not.

**Corollary 20.**  $\Theta$ -episturmian word is  $\Theta$ -rich if and only if  $\mathcal{P}_\Theta(1) + \mathcal{P}_\Theta(2) = \Delta\mathcal{C}(1) + 1$ .

### 5.1. Examples

This section contains examples of infinite words with properties mentioned above.

**Example 21.** Let  $\pi$  denote a morphism  $\{0, 1\}^* \mapsto \{a, a', c\}^*$ :

$$\pi : \begin{cases} 0 \mapsto aa' \\ 1 \mapsto aa'c. \end{cases}$$

Denote  $\Theta$  an involutory antimorphism defined as

$$\Theta : a \mapsto a', a' \mapsto a, c \mapsto c.$$

Then the infinite word  $\mathbf{u} = \pi(\mathbf{v})$ , where  $\mathbf{v}$  is a Sturmian word, has the following properties:

- (1) is  $\Theta$ -episturmian,
- (2) satisfies the equality  $\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n + 1) = \Delta\mathcal{C}(n) + 2$  for  $n \geq 1$ ,
- (3) is  $\Theta$ -rich.

As  $\mathbf{v}$  is Sturmian, it is clear that  $\mathbf{u}$  is aperiodic. Furthermore if we enumerate  $\mathcal{L}_2(\mathbf{u})$  we see that  $\Delta\mathcal{C}(1) = 1$ . These two facts imply that for  $n \geq 1$  we have  $\Delta\mathcal{C}(n) = 1$ . As the language  $\mathcal{L}(\mathbf{v})$  is closed under reversal, the language  $\mathcal{L}(\mathbf{u})$  is closed under  $\Theta$ .

To prove the rest we can either apply directly [Corollary 20](#) or enumerate  $\Theta$ -palindromic complexity.

In the following examples the mapping  $\Theta$  is considered the same as in [Example 21](#).

**Example 22.** Let  $\pi$  be the following morphism

$$\pi : 0 \mapsto a'cacc, 1 \mapsto a'cac.$$

We set  $\mathbf{u} = \pi(\mathbf{v})$  where  $\mathbf{v}$  is a Sturmian word over  $\{0, 1\}$ . Then  $\mathbf{u}$  is  $\Theta$ -episturmian and not  $\Theta$ -rich.

The proof is similar as in [Example 21](#). From  $\mathbf{v}$  we can deduce that the word  $\mathbf{u}$  is aperiodic, uniformly recurrent and contains infinitely many  $\Theta$ -palindromes. The last two properties imply that  $\mathcal{L}(\mathbf{u})$  is closed under  $\Theta$ . As  $\Delta\mathcal{C}(2) = 1$  and there is one special factor of length 1 – the factor  $c$  – it is clear that  $\mathbf{u}$  is  $\Theta$ -episturmian.

$\mathbf{u}$  is not  $\Theta$ -rich: factors  $ca'c$  and  $cac$  do not have a unioccurrent  $\Theta$ -palindromic suffix.

The next example is a  $\Theta$ -rich word which is not  $\Theta$ -episturmian.

**Example 23.** Set  $\pi$  as follows

$$\pi : 0 \mapsto aa', 1 \mapsto acca'.$$

Let  $\mathbf{v}$  be a Sturmian word over  $\{0, 1\}$ . Then the word  $\pi(\mathbf{v})$  is  $\Theta$ -rich and is not  $\Theta$ -episturmian. The proof is left to the reader.

The following two examples concern periodic infinite words.

**Example 24.** The word  $(caca')^\omega$  is periodic but not  $\Theta$ -rich. The factors  $ca'c$  and  $cac$  do not have a unioccurrent  $\Theta$ -palindromic suffix as before.

**Example 25.** The word  $(ccaa')^\omega$  is periodic and  $\Theta$ -rich. It follows from the fact that we can find a unioccurrent  $\Theta$ -palindromic suffix for each prefix of this word.

## 6. Final remarks

A wider class of infinite words related to  $\Theta$ -palindromes has been studied in [7]. The class is constructed using the *right  $\Theta$ -palindromic closure operator*. To any word  $w$  in  $\mathcal{A}^*$  it associates the word  $w^\oplus$  which is the shortest  $\Theta$ -palindromic word having  $w$  as prefix. It is clear that if  $s$  denotes the longest  $\Theta$ -palindromic suffix of  $w$  then if we denote  $w = ps$  we have  $w^\oplus = ps\Theta(p)$ .

Let  $u_0$  be a fixed finite word called *seed*. Let  $\Psi : \mathcal{A}^* \rightarrow \mathcal{A}^*$  be a map defined as follows

$$\begin{aligned} \Psi(\varepsilon) &= u_0, \\ \Psi(ux) &= (\Psi(u)x)^\oplus \text{ for } u \in \mathcal{A}^* \text{ and } x \in \mathcal{A}. \end{aligned}$$

For a given infinite word  $t$  we can define  $\Psi(t)$  as

$$\Psi(t) = \lim_{n \rightarrow +\infty} \Psi(\text{pref}_n(t))$$

where  $\text{pref}_n$  stands for the prefix of length  $n$ . The word  $t$  is called the *directive word* of  $\Psi(t)$  and is denoted  $\Delta(\Psi(t))$ . The word  $\Psi(t)$  is called  *$\Theta$ -standard word with seed* if  $u_0 \neq \varepsilon$  otherwise it is called just  $\Theta$ -standard.  $\Theta$ -episturmian words are included in  $\Theta$ -standard words with seed (exact condition and details are given in [7]).

As the definition of  $\Theta$ -standard word is constructive, a rough and direct condition to decide whether a  $\Theta$ -standard word is  $\Theta$ -rich is easy to derive. If the word  $\Psi(\text{pref}_n(t))t_n$  is  $\Theta$ -rich for all  $n \geq 0$  (with  $t = t_0t_1t_2\dots$ ), then  $\Psi(t)$  is  $\Theta$ -rich.

The following proposition connects the class of  $\Theta$ -standard words with seed with the class of  $\Theta$ -rich words.

**Proposition 26** (*Proposition 4.7, [7]*). *Let  $\mathbf{u}$  be a  $\Theta$ -standard word with seed. Then there exists an integer  $M$  such that any prefix  $p$  of  $\mathbf{u}$  of length greater than  $M$  has a unioccurrent  $\Theta$ -palindromic suffix.*

The integer  $M$  is explained in detail in the paper. It is clear that if  $M = 1$ , then  $\mathbf{u}$  is  $\Theta$ -rich. However  $M = 1$  implies some restrictions on  $\Theta$  and a criterion whether a  $\Theta$ -standard word with seed is  $\Theta$ -rich remains an open question.

## Acknowledgements

I would like to thank Edita Pelantová and L'ubomíra Balková for their careful reviewing and fruitful suggestions. I acknowledge financial support by the Czech Science Foundation grant no. 201/09/0584, by the grant no. MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic and by the grant no. SGS10/085OHK4/1T/14 of the Grant Agency of the Czech Technical University in Prague.

## References

- [1] Petr Ambrož, Christiane Frougny, Zuzana Masáková, Edita Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers, *Ann. Inst. Fourier (Grenoble)* 56 (2006) 2131–2160.
- [2] Vyoma Anne, Luca Q. Zamboni, Ioana Zorca, Palindromes and pseudo-palindromes in episturmian and pseudo-episturmian infinite words, in: S. Brlek, C. Reutenauer (Eds.), *Words 2005*, in: *Publications du LACIM*, vol. 36, 2005, pp. 91–100.
- [3] Peter Baláži, Zuzana Masáková, Edita Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words, *Theoret. Comput. Sci.* 380 (3) (2007) 266–275.
- [4] L'ubomíra Balková, Edita Pelantová, Štěpán Starosta, Sturmian jungle (or garden?) on multiliteral alphabets, *RAIRO - Theor. Inform. Appl.* (in press).
- [5] Jean Berstel, Sturmian and episturmian words: a survey of some recent results, in: *CAI'07: Proceedings of the 2nd International Conference on Algebraic Informatics (Berlin, Heidelberg)*, in: *Lecture Notes in Computer Science*, vol. 4728, Springer-Verlag, 2007, pp. 23–47.
- [6] Srećko Brlek, Sylvie Hamel, Maurice Nivat, Christophe Reutenauer, On the palindromic complexity of infinite words, *Internat. J. Found. Comput. Sci.* 15 (2) (2004) 293–306.
- [7] Michelangelo Bucci, Aldo de Luca, Alessandro De Luca, Luca Q. Zamboni, On different generalizations of episturmian words, *Theoret. Comput. Sci.* 393 (1–3) (2008) 23–36.
- [8] Michelangelo Bucci, Aldo de Luca, Alessandro De Luca, Luca Q. Zamboni, On theta-episturmian words, *European J. Combin.* 30 (2) (2009) 473–479.
- [9] Michelangelo Bucci, Alessandro De Luca, Amy Glen, Luca Q. Zamboni, A connection between palindromic and factor complexity using return words, *Adv. Appl. Math.* 42 (1) (2009) 60–74.
- [10] Michelangelo Bucci, Alessandro De Luca, Amy Glen, Luca Q. Zamboni, A new characteristic property of rich words, *Theoret. Comput. Sci.* 410 (30–32) (2009) 2860–2863.
- [11] Xavier Droubay, Jacques Justin, Giuseppe Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* 255 (1–2) (2001) 539–553.
- [12] Xavier Droubay, Giuseppe Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* 223 (1–2) (1999) 73–85.
- [13] Amy Glen, Jacques Justin, Steve Widmer, Luca Q. Zamboni, Palindromic richness, *European J. Combin.* 30 (2) (2009) 510–531.
- [14] Glen Amy, Justin Jacques, Episturmian words: a survey, *RAIRO - Theor. Inform. Appl.* 43 (3) (2009) 403–442.